

LMI Training Institute  
June 16-20, 2003

Prepared by the Staff of the LEHD Program

June 2, 2003

## Chapter 4

# QWI Disclosure Limitation

### 4.1 Disclosure-proofing the QWI

Confidentiality protection is an essential feature of any statistical system. The providers of the unemployment insurance wage records, the ES-202 establishment records, and all of the confidential Title 13 data at the Census Bureau require that the information released in the QWI not reveal the values of any individual's or business's data. Disclosure-proofing is the technical name for the methods that we use to ensure that the underlying microdata remain confidential. In this subsection we explain the disclosure-proofing methodology and how to use the status flags that we provide for each QWI in every quarter. These status flags summarize the effects of the disclosure-proofing on the reported data item.

The QWI are estimates of a variety of work force measures constructed at very detailed geographical and industrial levels. In order to protect the confidentiality of the reports on individual workers and specific businesses, we must insure that none of the indicators can be used to deduce the value of a measure for an identifiable person or establishment. We accomplish this protection using two related techniques. First, every business establishment in the QWI data bases is assigned a fuzz factor. The fuzz factor, which is constant over time for a given business establishment, is used to distort all of the measures for that establishment that enter the computation of the QWIs. Since every data item for every business establishment in every quarter is distorted, the actual values of the underlying micro data are never used in the computation of any QWI. However, the statistical properties of the distortion factor are carefully controlled so that as we form the QWIs by combining the data from multiple business establishments the average distortion is reduced until it is, generally, negligible. Second, QWIs that measure employment counts based on one or two individuals or based on one or two businesses are suppressed.

The QWI disclosure-proofing methodology has undergone rigorous testing to determine the extent to which it affects the usefulness of the reported data. Our research has demonstrated that these methods produce far more useful QWIs than the competing methodology, which suppresses all the data items when the number of individuals or businesses in a particular geography-industry combination is small. In addition to allowing the Census Bureau to release the QWIs at detailed levels of geography and industry, our disclosure-proofing methodology also allows the release of detailed age and sex demographics within a given geography-industry cell. The competing confidentiality protection method would have suppressed virtually all of this detail. Once we have completed the disclosure-proofing of a particular data item, we assign the item a status flag. The value of the status flag for a particular QWI varies from quarter to quarter depending upon the composition of the work force and establishments that contribute to its value.

The status flag is a very important part of the QWI data. The Census Bureau requires that QWI data items always be reported with their associated status flags. For this reason, we explain how to use the disclosure status variable. The methodology is explained in detail in the Excel workbook "Disclosure\_control\_table\_public\_use-20030531.xls" included in the electronic version of this documentation.

When the disclosure-proofing methodology has not affected the value of particular data item for a given quarter, we give that quarterly data item a disclosure status value of 1 (one). For example, in the public use file beginning of quarter employment is called BeginEmp. Its associated disclosure status variable is sBeginEmp. For a given state, year, quarter, geography, industry, sex, and age, if sBeginEmp=1, then the disclosure methodology has had no appreciable effect on the value of BeginEmp in that given set of categories. Other variables in the QWI files work the same way

and have their own disclosure status variables. For QWI data items, when the quarterly disclosure status variable has the value of 1 you can be confident that the confidentiality protection has not materially affected the value of the indicator you are using. Most QWIs have a status flag of 1.

When the disclosure-proofing has resulted in a significant distortion of the underlying data item, we give that quarterly data item a status value of 9 (nine). When you encounter a QWI value whose status flag is 9, you should inform your users that the value has been significantly distorted to protect the confidentiality of the underlying micro data. Using *BeginEmp* as an example again, when *sBeginEmp*=9 the value of *BeginEmp* for a given state, year, quarter, geography, industry, sex, and age has been significantly distorted to protect the confidentiality of the underlying microdata. You should use this value with some caution; however, we strongly recommend that you not suppress these values. Our research shows that even when the indicator that we release is significantly distorted its time series properties are not significantly distorted. Changes in the QWIs from quarter to quarter are reliable even if the status flag for a particular data item indicates that it has been significantly distorted.

We now consider some other possibilities for the disclosure status variable. These other possibilities depend upon how many employers are operating in a given state, year, quarter, geography, industry category. They also depend upon how many employees these employers have, in aggregate, in a given sex, age category. Lastly, they depend upon whether the indicator we are considering is a count, sum, or ratio.

Consider first the case where there are some employers operating in a given state, year, quarter, geography, industry category. When the QWI is an employment count (beginning-of-quarter employment, end-of-quarter employment, full-quarter employment, and the associated accession and separation measures) that is based upon one or two employees in a given sex, age category, we assign a quarterly disclosure status variable a value of 2 (two). When the disclosure status value is 2, the QWI is not reported. We suppress the value of the count indicator when there are one or two employees in the category to protect the identity of those employees in the microdata. For sums and ratios, our regular confidentiality protections are sufficient. You will never see a disclosure status value of 2 for sums (e.g., payroll) or ratios (e.g., average earnings of full quarter employees).

The next possibility is that there are only one or two employers operating in a given state, year, quarter, geography, industry category. We believe that our disclosure protection methodology works well for this case; however, we have been asked to study it further. For this case, the disclosure status variable takes a value of 3 (three). Employment count data are not reported when the disclosure status variable has a value of 3. For sums and ratios, our disclosure protection methodology has been shown to provide sufficient protection in this case. You will never see a disclosure status value of 3 for sums or ratios.

The status variable can take two additional values. These values explain why there are no data available for certain variables in certain state, year, quarter, geography, industry categories.

When the quarterly unemployment insurance wage record information required to compute a particular QWI is not available, we assign the data item a status flag of -1 (minus one). Status flags of -1 occur at the beginning of a state's historical data for QWIs associated with beginning-of-quarter employment, accessions, hires, recalls, job creations and job destructions. Until we have accumulated enough quarters of initial UI wage record data, we simply cannot compute some of these indicators. Status flags of -1 also occur in the most recent quarters for QWIs associated with end-of-quarter employment separations, job creations and job destructions. The most recent values of these indicators require additional quarters of information beyond the most recently received UI wage records.

It sometimes occurs that there are no employers operating in a particular state, year, quarter, geography, industry category but there were employers operating in that state, geography, industry during other quarters. When this occurs the disclosure status variable takes the value -2 for all variables in the quarters in which there were no employers operating in the category. No values of the indicators are available for these quarters. We left the observations in public use files to facilitate converting between tabular and time series analyses. If a particular state, geography, industry category ever has a positive number of employers, the QWIs provide a complete historical time series for this category. When the disclosure status value is -2, the "not available" value can be treated as a zero for the time series analysis of counts and sums.

## 4.2 Summary

The table below summarizes the meaning of the values of the disclosure status variable.

<b>Value</b>	<b>Meaning of the disclosure status variable</b>
-2	No employers operate in this category during this quarter
-1	Required historical or future wage record data are not available
0	No employment in this age, sex category
1	The value is not significantly affected by the disclosure protections.
2	There are less than 3 employees in this category
3	There are less than 3 employers in this category
9	The value has been significantly distorted to protect the confidentiality

### QWI Disclosure Proofing Algorithm

let qfirst = the first available quarter of QWI data

let qlast = the last available quarter of QWI data

let iq = the current quarter

let nl = allowable noise limit

if disclosure\_type = count then do

    if (iq < qfirst + QWI\_historical\_quarters\_needed) or (iq > qlast - QWI\_future\_quarters\_needed) then disclosure\_status = -1

    else if (Employee\_disclosure\_stem\_1 = 0) and (Employee\_disclosure\_stem\_2 = 0) and (Employee\_disclosure\_stem\_3 = 0) then disclosure\_status = -1

    else if (Employee\_disclosure\_stem\_1 < 3) or (Employee\_disclosure\_stem\_2 < 3) or (Employee\_disclosure\_stem\_3 < 3) then disclosure\_status = -1

    else if (Employer\_disclosure\_stem\_1 < 3) or (Employer\_disclosure\_stem\_2 < 3) or (Employer\_disclosure\_stem\_3 < 3) then disclosure\_status = -1

    else if abs(np\_Variable\_stem) > nl then disclosure\_status = 9 /\* np = noise percentage \*/

    else disclosure\_status = 1

    if max(of all disclosure\_status variables) = missing then for all variables disclosure\_status = -2

end

else if disclosure\_type = ratio or disclosure\_type = sum then do

    if (iq < qfirst + QWI\_historical\_quarters\_needed) or (iq > qlast - QWI\_future\_quarters\_needed) then disclosure\_status = -1

    else if (Employee\_disclosure\_stem\_1 = 0) and (Employee\_disclosure\_stem\_2 = 0) and (Employee\_disclosure\_stem\_3 = 0) then disclosure\_status = -1

    else if (Employee\_disclosure\_stem\_1 < 3) or (Employee\_disclosure\_stem\_2 < 3) or (Employee\_disclosure\_stem\_3 < 3) then disclosure\_status = -1

    else if (Employer\_disclosure\_stem\_1 < 3) or (Employer\_disclosure\_stem\_2 < 3) or (Employer\_disclosure\_stem\_3 < 3) then disclosure\_status = -1

    else if abs(np\_Variable\_stem) > nl then disclosure\_status = 9 /\* np = noise percentage \*/

    else disclosure\_status = 1

    if max(of all disclosure\_status variables) = missing then for all variables disclosure\_status = -2

end

else do

    traps errors in disclosure\_control\_values

end

codes

-2 there are no establishments in this state, year, quarter, geography, industry category for this quarter (but there are data for other quarters)

-1 unable to compute this estimate because historical data are not available or future quarters are required

0 there is no employment in this cell, but there are establishments in the category (OK to disclose a 0)

1 OK

2 less than 3 employees (value suppressed in publications)

3 less than 3 employers (value suppressed in publications)

9 data significantly distorted

Variable_stem	Employee_disclosure_stem_1	Employee_disclosure_stem_2	Employee_disclosure_stem_3	Employer_disclosure_stem_1	Employer_disclosure_stem_2	Employer_disclosure_stem_3	Disclosure_type	QWI_historical_quarters_needed	QWI_future_quarters_needed
B	B	B	B	n_B	n_B	n_B	count	1	0
E	E	E	E	n_E	n_E	n_E	count	0	1
JF	B	E	E	n_B	n_E	n_E	count	1	1
JC	B	E	E	n_B	n_E	n_E	count	1	1
JD	B	E	E	n_B	n_E	n_E	count	1	1
A	A	A	A	n_A	n_A	n_A	count	1	0
H	H	H	H	n_H	n_H	n_H	count	4	0
R	R	R	R	n_R	n_R	n_R	count	4	0
S	S	S	S	n_S	n_S	n_S	count	0	1
F	F	F	F	n_F	n_F	n_F	count	1	1
FJF	F	F	F	n_F	n_F	n_F	count	2	1
FJC	F	F	F	n_F	n_F	n_F	count	2	1
FJD	F	F	F	n_F	n_F	n_F	count	2	1
FT	F	FA	FS	n_F	n_FA	n_FS	ratio	2	1
FA	FA	FA	FA	n_FA	n_FA	n_FA	count	2	1
H3	H3	H3	H3	n_H3	n_H3	n_H3	count	5	1
FS	FS	FS	FS	n_FS	n_FS	n_FS	count	2	1
Z_W2	E	E	E	n_E	n_E	n_E	ratio	0	1
Z_W3	F	F	F	n_F	n_F	n_F	ratio	1	1
Z_WFA	FA	FA	FA	n_FA	n_FA	n_FA	ratio	2	1
Z_dWA	A	A	A	n_A	n_A	n_A	ratio	1	0
Z_NA	A	A	A	n_A	n_A	n_A	ratio	4	0
Z_NH	H	R	R	n_H	n_R	n_R	ratio	4	0
Z_NR	H	R	R	n_H	n_R	n_R	ratio	4	0
Z_WFS	FS	FS	FS	n_FS	n_FS	n_FS	ratio	2	2
Z_dWS	S	S	S	n_S	n_S	n_S	ratio	0	1
Z_NS	S	S	S	n_S	n_S	n_S	ratio	0	4
Z_WH3	H3	H3	H3	n_H3	n_H3	n_H3	ratio	5	1
W1	M	M	M	n_M	n_M	n_M	sum	0	0
M	M	M	M	n_M	n_M	n_M	count	0	0
bdot	bdot	bdot	bdot	n_bdot	n_bdot	n_bdot	count	1	0
Z_inwb	bdot	bdot	bdot	n_bdot	n_bdot	n_bdot	ratio	1	0
Z_inwblg	bdot	bdot	bdot	n_bdot	n_bdot	n_bdot	ratio	1	0
Z_inwb2	bdot	bdot	bdot	n_bdot	n_bdot	n_bdot	ratio	1	0
Z_inwb_inwblg	bdot	bdot	bdot	n_bdot	n_bdot	n_bdot	ratio	1	0
Z_inwblg2	bdot	bdot	bdot	n_bdot	n_bdot	n_bdot	ratio	1	0
edot	edot	edot	edot	n_edot	n_edot	n_edot	count	0	1
Z_inwe	edot	edot	edot	n_edot	n_edot	n_edot	ratio	0	1
Z_inweld	edot	edot	edot	n_edot	n_edot	n_edot	ratio	0	1
Z_inwe2	edot	edot	edot	n_edot	n_edot	n_edot	ratio	0	1
Z_inwe_inweld	edot	edot	edot	n_edot	n_edot	n_edot	ratio	0	1
Z_inweld2	edot	edot	edot	n_edot	n_edot	n_edot	ratio	0	1
fdot	fdot	fdot	fdot	n_fdot	n_fdot	n_fdot	count	1	1
Z_inwf	fdot	fdot	fdot	n_fdot	n_fdot	n_fdot	ratio	1	1
Z_inwflid	fdot	fdot	fdot	n_fdot	n_fdot	n_fdot	ratio	1	1
Z_inwflg	fdot	fdot	fdot	n_fdot	n_fdot	n_fdot	ratio	1	1

Variable_stem	Employee_disclosure_stem_1	Employee_disclosure_stem_2	Employee_disclosure_stem_3	Employer_disclosure_stem_1	Employer_disclosure_stem_2	Employer_disclosure_stem_3	Disclosure_type	QWI_historical_quarters_needed	QWI_future_quarters_needed
Z_inwf2	fdot	fdot	fdot	n_fdot	n_fdot	n_fdot	ratio	1	1
Z_inwfid2	fdot	fdot	fdot	n_fdot	n_fdot	n_fdot	ratio	1	1
Z_inwflg2	fdot	fdot	fdot	n_fdot	n_fdot	n_fdot	ratio	1	1
Z_inwf_inwfid	fdot	fdot	fdot	n_fdot	n_fdot	n_fdot	ratio	1	1
Z_inwf_inwflg	fdot	fdot	fdot	n_fdot	n_fdot	n_fdot	ratio	1	1
Z_inwfid_inwflg	fdot	fdot	fdot	n_fdot	n_fdot	n_fdot	ratio	1	1

**QWI AFF Level Variables**

if status\_F ^= -1 and F>0 then FT =ROUND(0.5\*(FA + FS)/F,3)  
else FT=(SAS missing) and status\_FT=-1

Note: full quarter turnover must be calculated from the AFF. Each of the statistics (FA, FS, and F is taken from the correct age x sex group. The statistic is undefined if F=0, so we must reset its disclosure status flag to -1. The result is a number usually less than 1 (although it can be greater than 1), which should be rounded to 3 significant decimal places. Note that the disclosure proofing is based on F, FA and FS because all three items are also contained in the public use file, so the statistic is dependent on the disclosure status of all three.